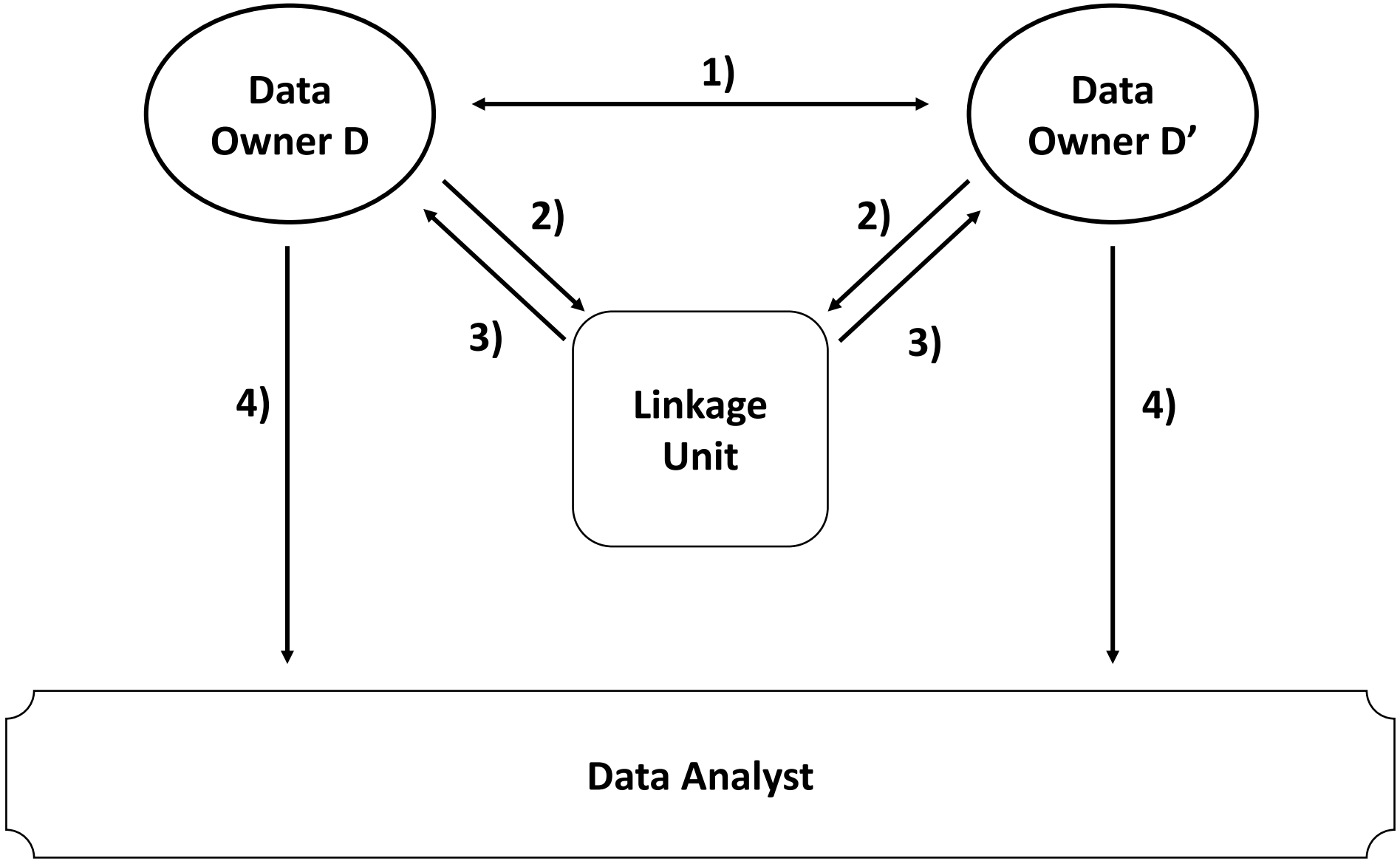
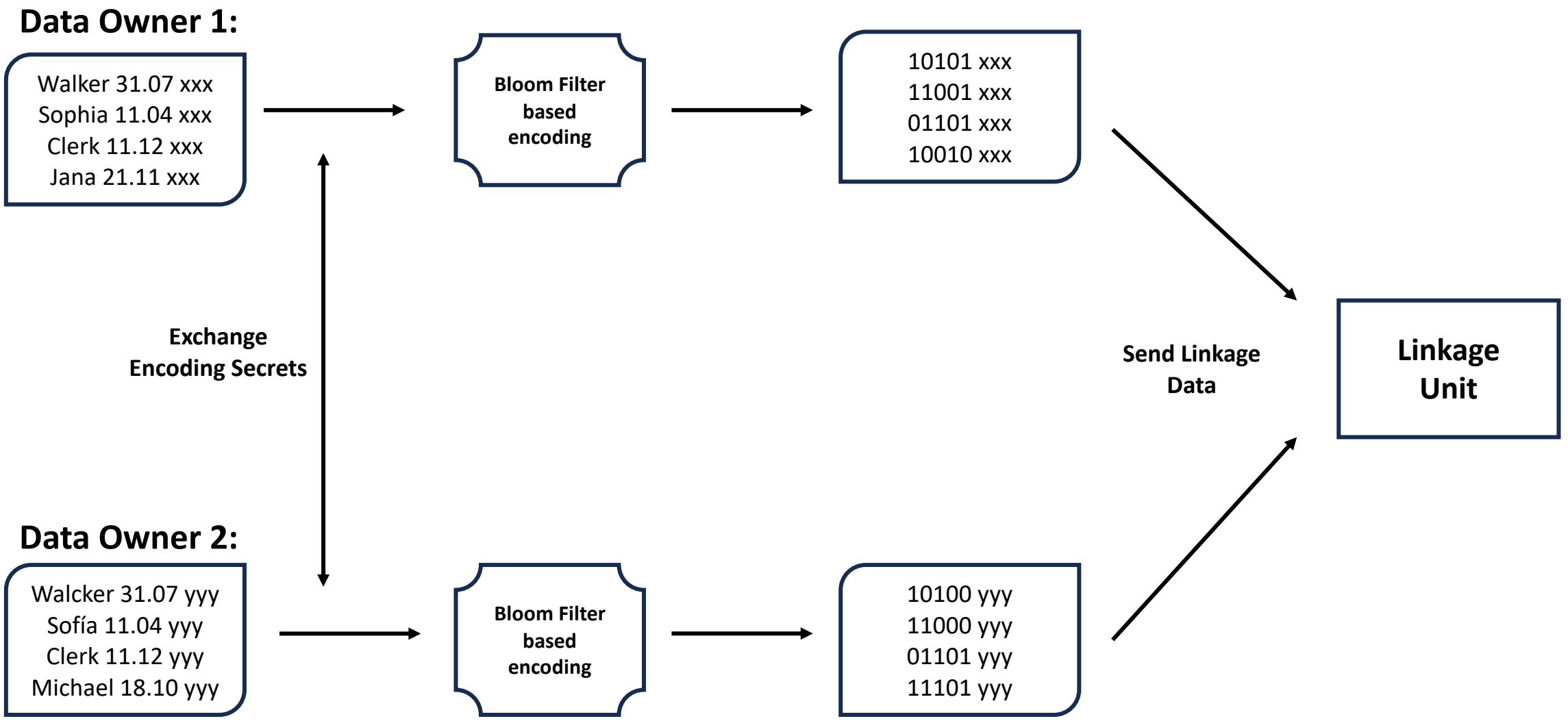
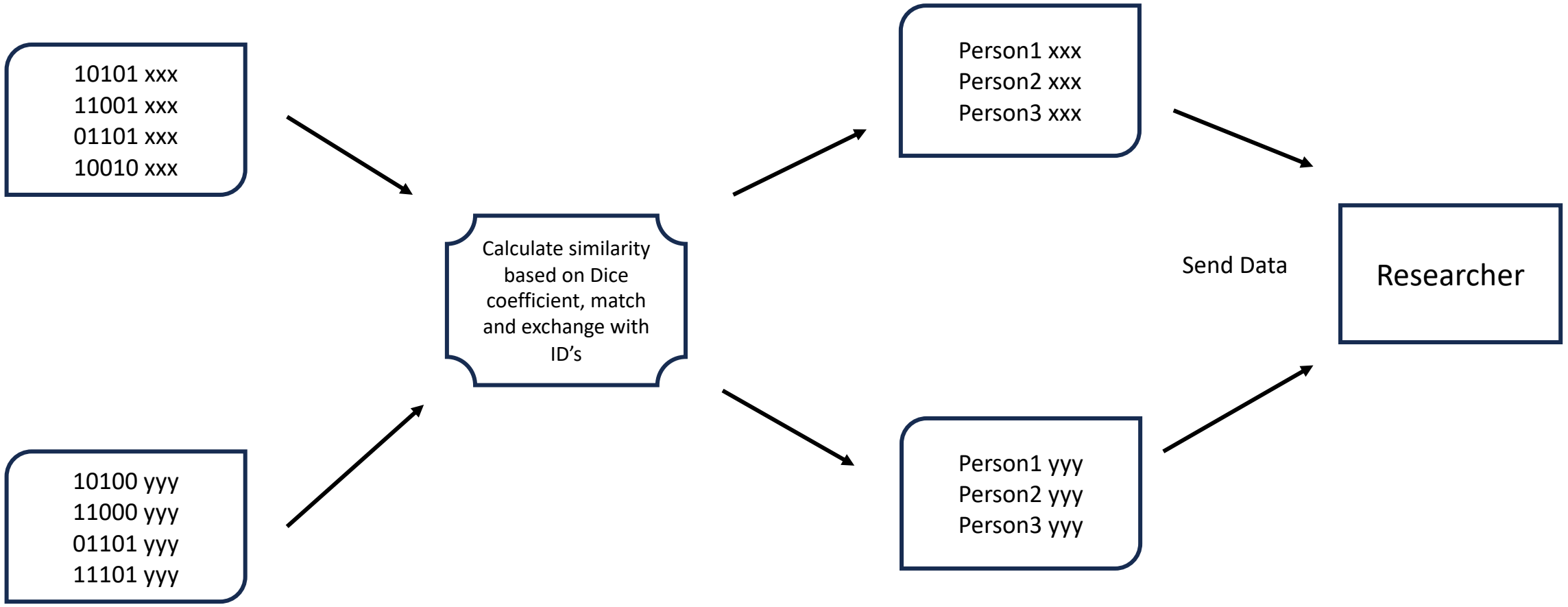
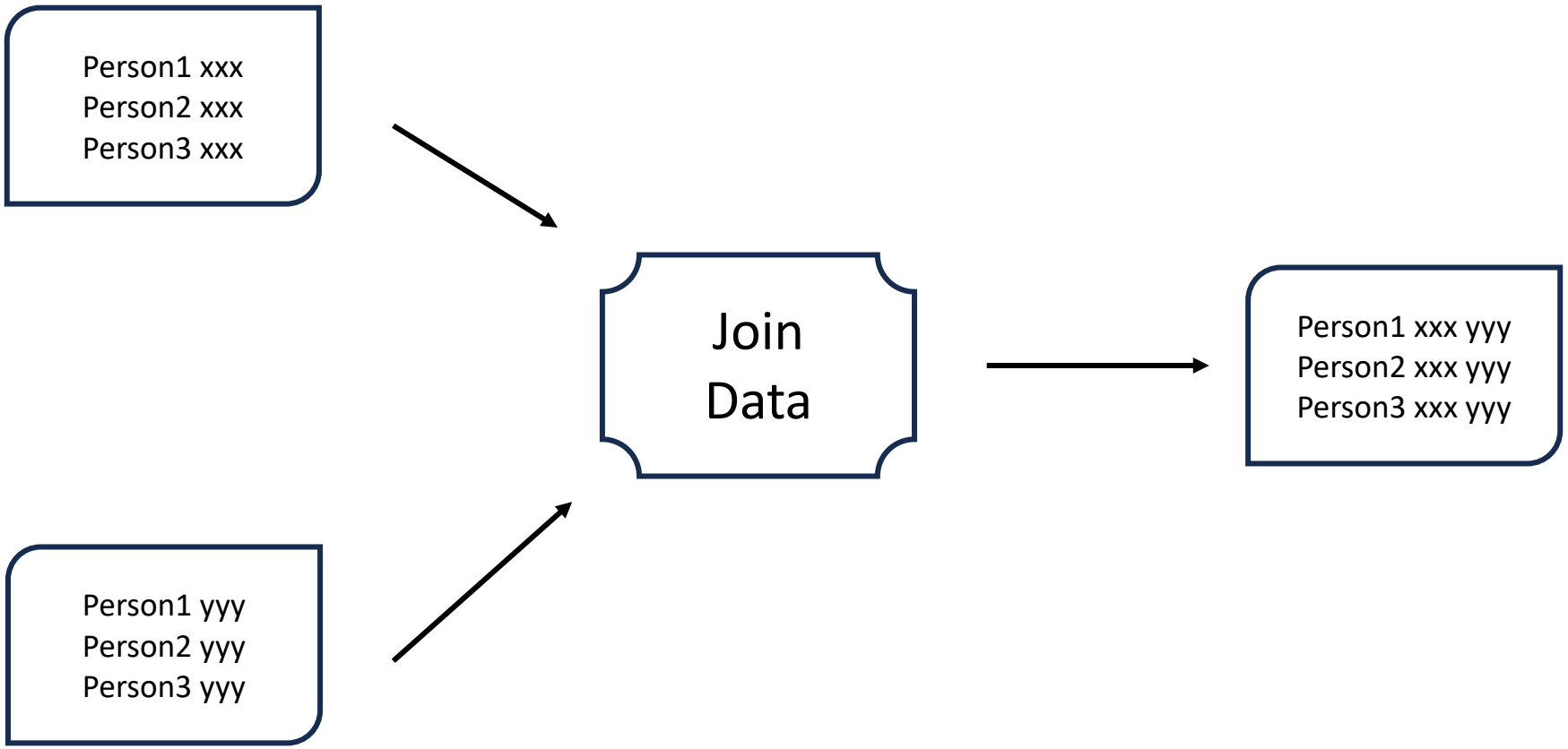

Data
Owner D
1)
Data
Owner D'
2)
2)
3)
3)
4)
Linkage
Unit
4)
Data Analyst

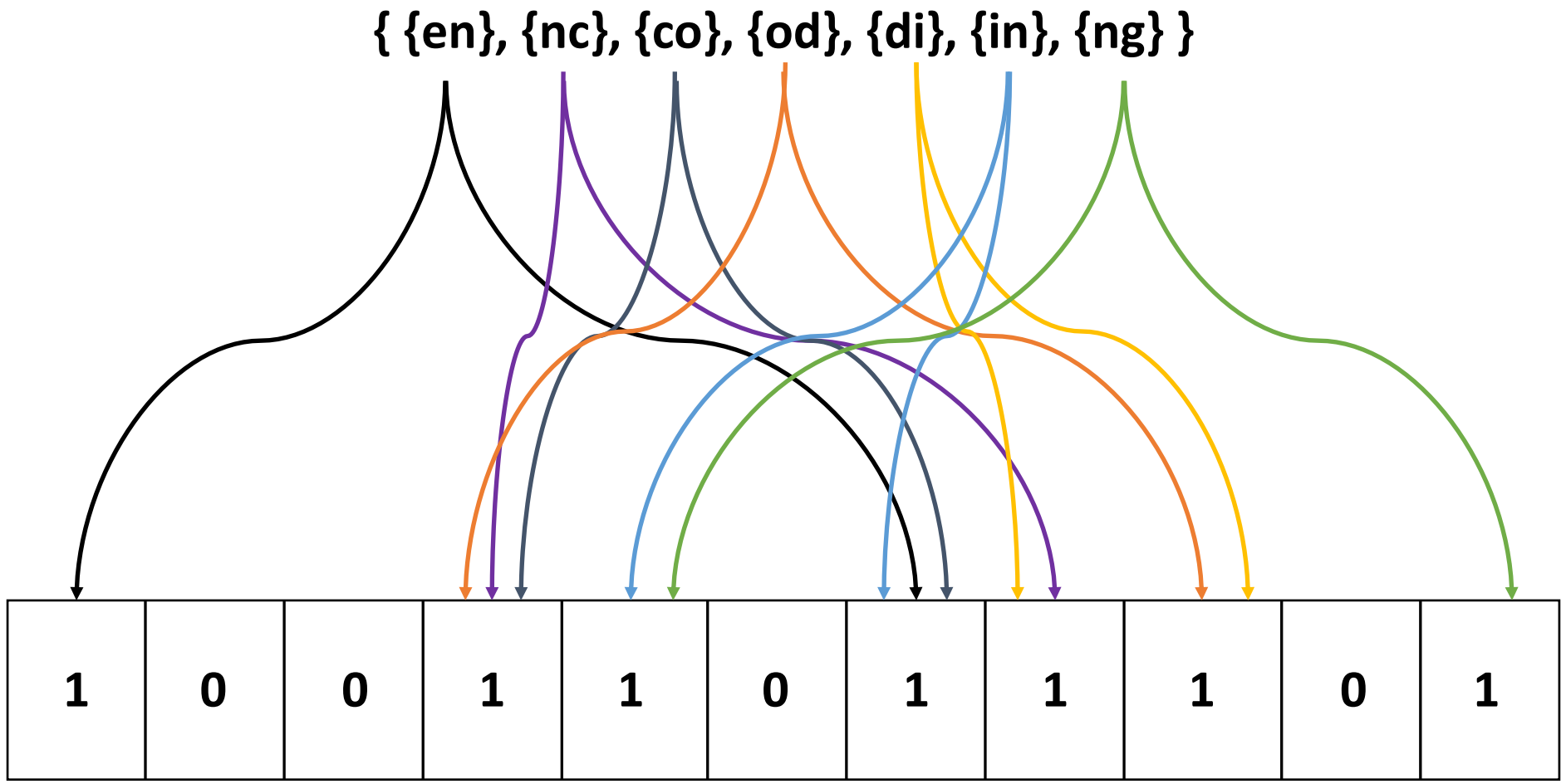

Data Owner 1:
Walker 31.07 xxx
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 21.11 xxx
Bloom Filter based encoding
10101 xxx
11001 xxx
01101 xxx
10010 xxx
Exchange Encoding Secrets
Send Linkage Data
Linkage Unit
Data Owner 2:
Walcker 31.07 yyy
Sofía 11.04 yyy
Clerk 11.12 yyy
Michael 18.10 yyy
Bloom Filter based encoding
10100 yyy
11000 yyy
01101 yyy
11101 yyy

Linkage Unit:



Researcher:


Person1 xxx
Person2 xxx
Person3 xxx
Person1 yyy
Person2 yyy
Person3 yyy
Join
Data
Person1 xxx yyy
Person2 xxx yyy
Person3 xxx yyy


{ {en}, {nc}, {co}, {od}, {di}, {in}, {ng} }
1
0
0
1
1
0
1
1
1
0
1

Element	S	S’
1	1	1
2	0	1
3	1	1
4	1	0

π_1

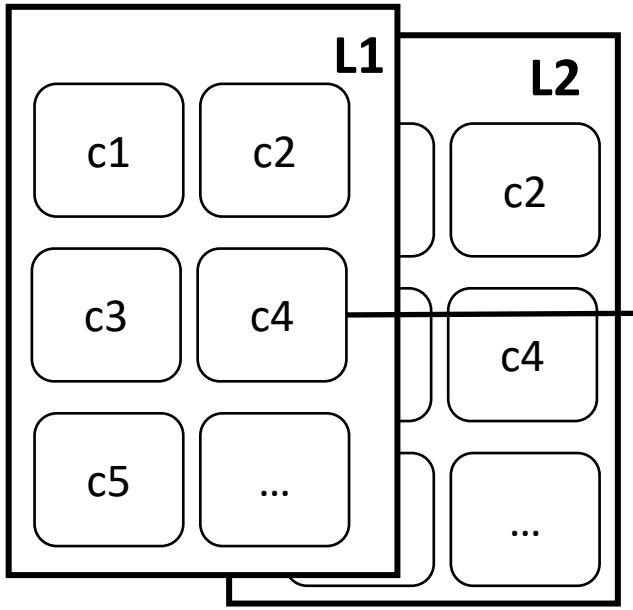
Element	S	S’
4	1	0
3	1	1
2	0	1
1	1	1

$[\ 1(\pi_1(S_1)=\pi_1(S'_1) + 1(\pi_2(S_1)=\pi_2(S'_1)\ ] / 2$
$= (1 * 0 + 1 * 1) / 2$
$= 1/2$

π_2

Element	S	S’
3	1	1
4	1	0
2	0	1
1	1	1

Initialization



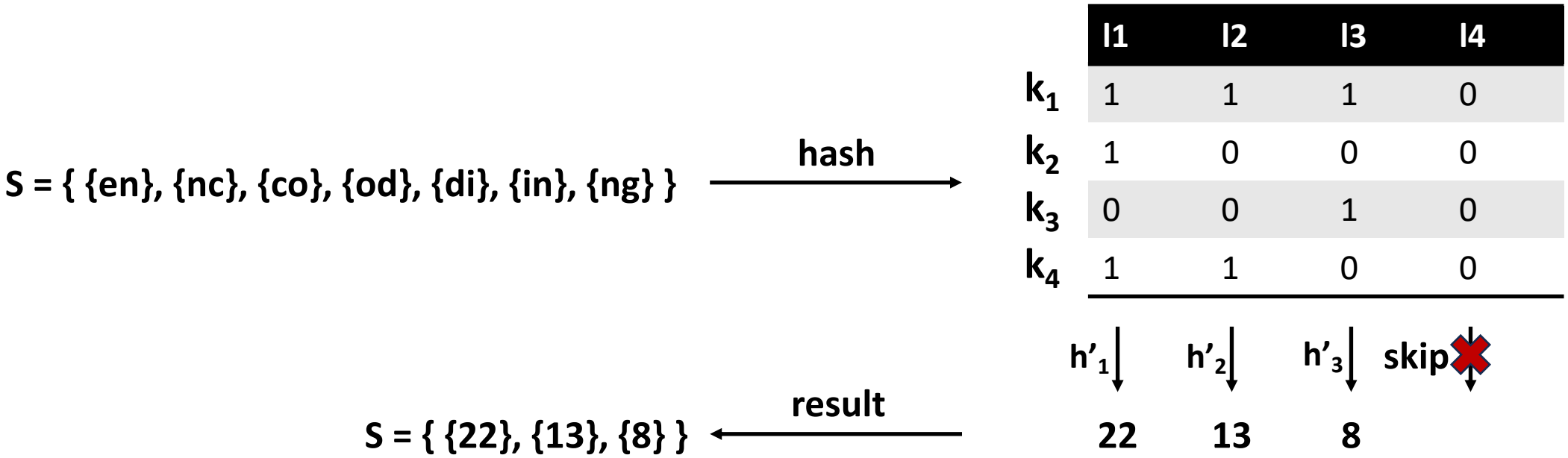
Key	Bit string
10	101001
11	011001
01	100100
11	111000

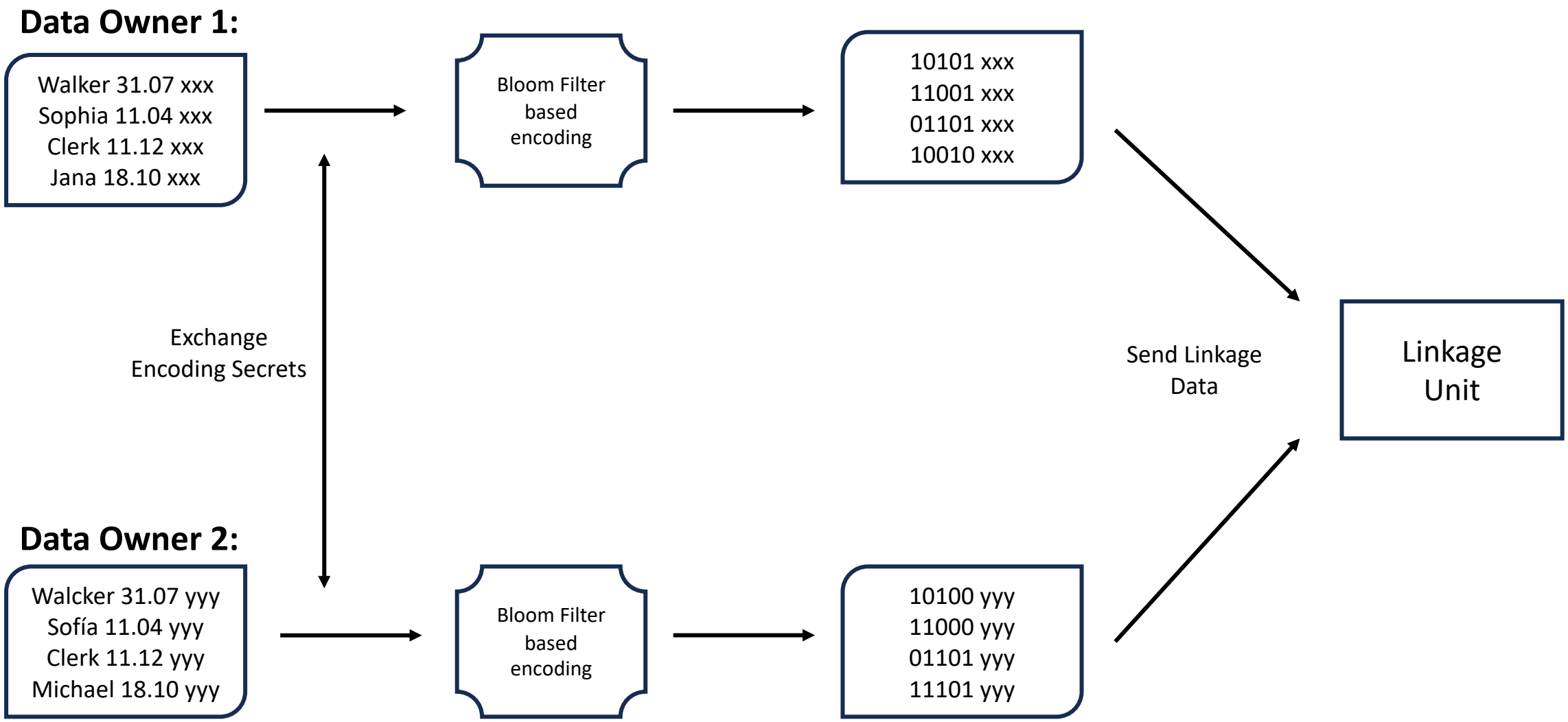
S = { {en}, {nc}, {co}, {od}, {di}, {in}, {ng} }

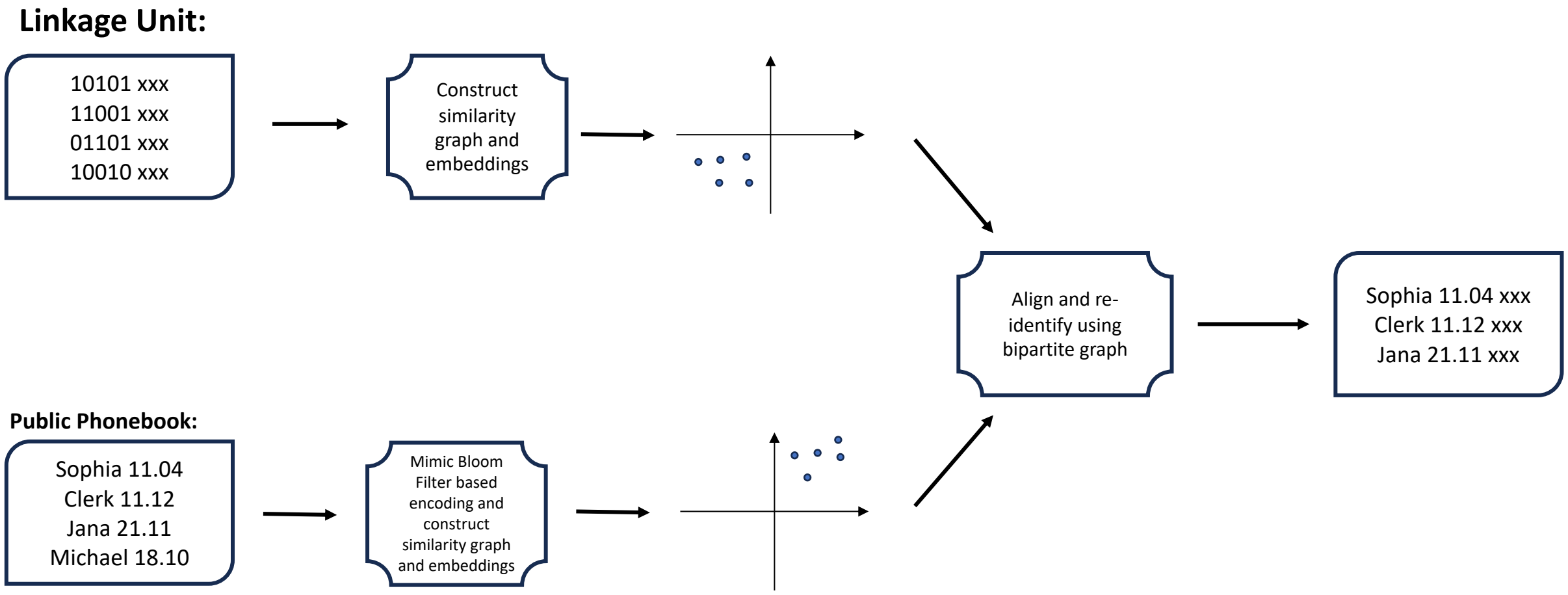
Hasing Process

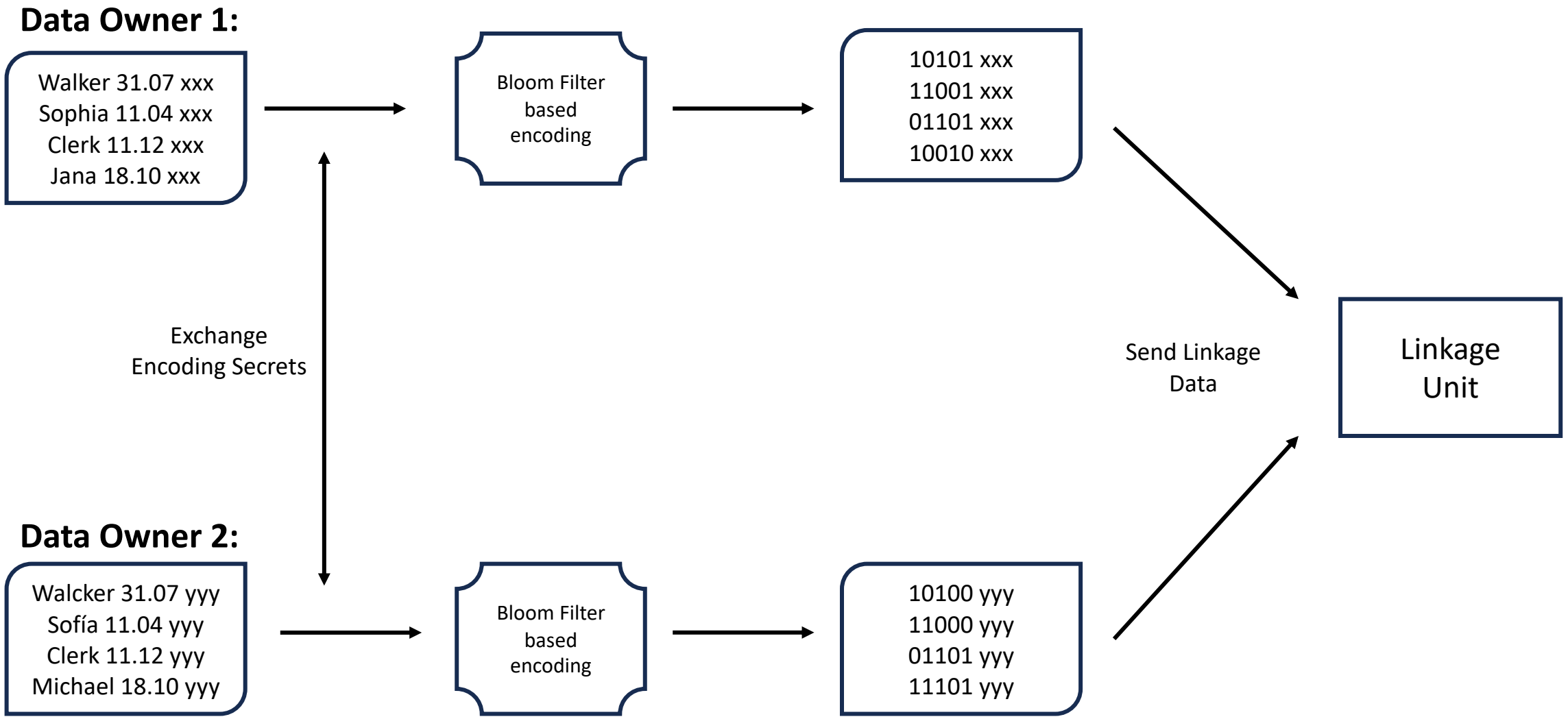
S_1 = {en} —h_4→ **10010111** —split→ **{ {10}, {01}, {01}, {11} }** —Look up→ **{ {101001}, {100100}, {100100}, {111000} }**

{ {101001}, {100100}, {100100}, {111000} } —XOR→ **010001**

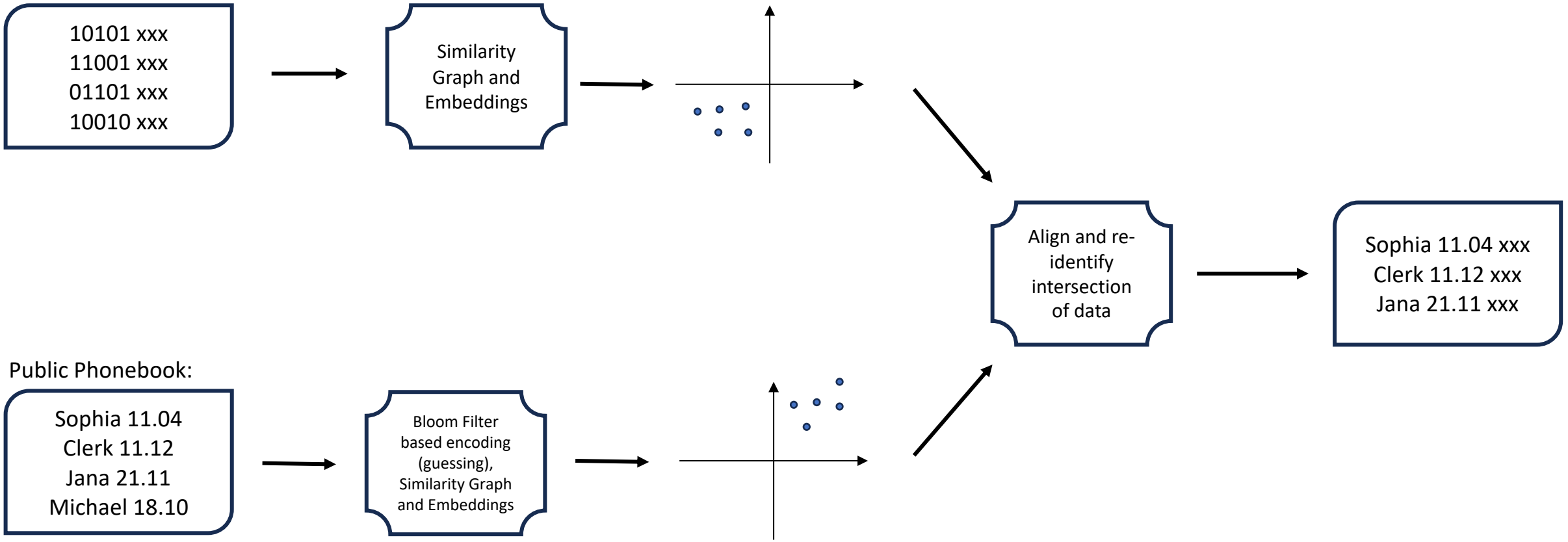

S = { {en}, {nc}, {co}, {od}, {di}, {in}, {ng} }
hash
l1 l2 l3 l4
k1 1 1 1 0
k2 1 0 0 0
k3 0 0 1 0
k4 1 1 0 0
h'1 h'2 h'3 skip
result
S = { {22}, {13}, {8} }
22 13 8


Data Owner 1:
Walker 31.07 xxx
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 18.10 xxx
Bloom Filter based encoding
10101 xxx
11001 xxx
01101 xxx
10010 xxx
Exchange Encoding Secrets
Send Linkage Data
Linkage Unit
Data Owner 2:
Walcker 31.07 yyy
Sofía 11.04 yyy
Clerk 11.12 yyy
Michael 18.10 yyy
Bloom Filter based encoding
10100 yyy
11000 yyy
01101 yyy
11101 yyy


Linkage Unit:
10101 xxx
11001 xxx
01101 xxx
10010 xxx
Construct similarity graph and embeddings
Public Phonebook:
Sophia 11.04
Clerk 11.12
Jana 21.11
Michael 18.10
Mimic Bloom Filter based encoding and construct similarity graph and embeddings
Align and re-identify using bipartite graph
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 21.11 xxx


Data Owner 1:
Walker 31.07 xxx
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 18.10 xxx
Bloom Filter based encoding
10101 xxx
11001 xxx
01101 xxx
10010 xxx
Exchange Encoding Secrets
Send Linkage Data
Linkage Unit
Data Owner 2:
Walcker 31.07 yyy
Sofía 11.04 yyy
Clerk 11.12 yyy
Michael 18.10 yyy
Bloom Filter based encoding
10100 yyy
11000 yyy
01101 yyy
11101 yyy

Linkage Unit:



Linkage Unit:

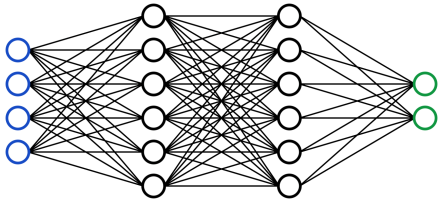
Known:

10101 **Sophia 11.04** xxx
11001 **Clerk 11.12** xxx
01101 **Jana 21.11** xxx

Train Neural Network

Unknown:

10010 xxx

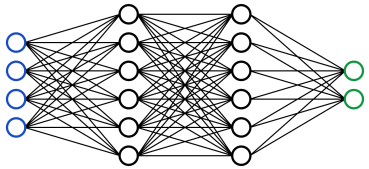


10010 **Walker 31.07** xxx

Data Extraction → Data Preparation → Model Training → Application to Encoded Data → Refinement and Reconstruction

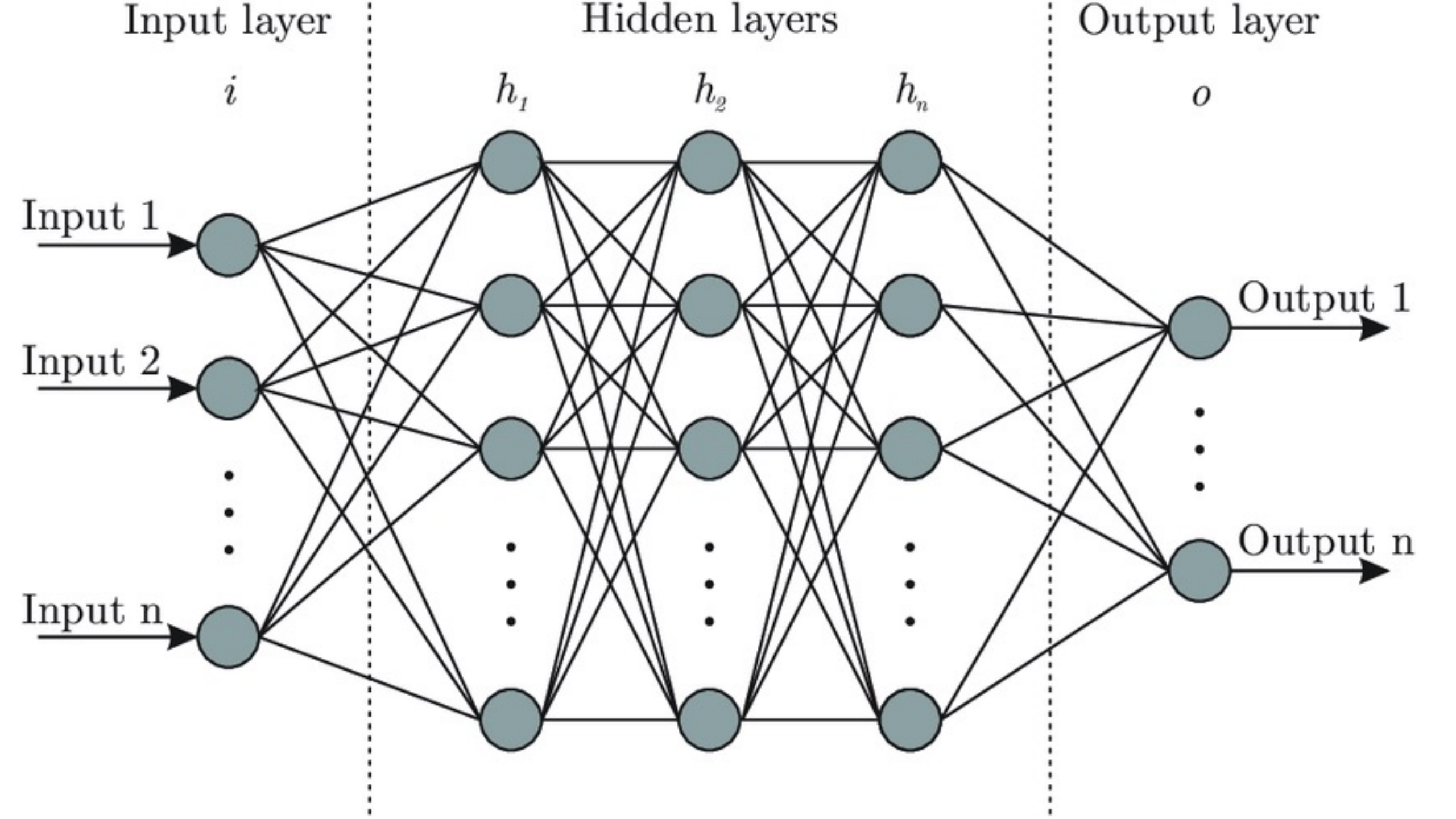
10101 Sophia 11.04 xxx
11001 Clerk 11.12 xxx
01101 Jana 21.11 xxx

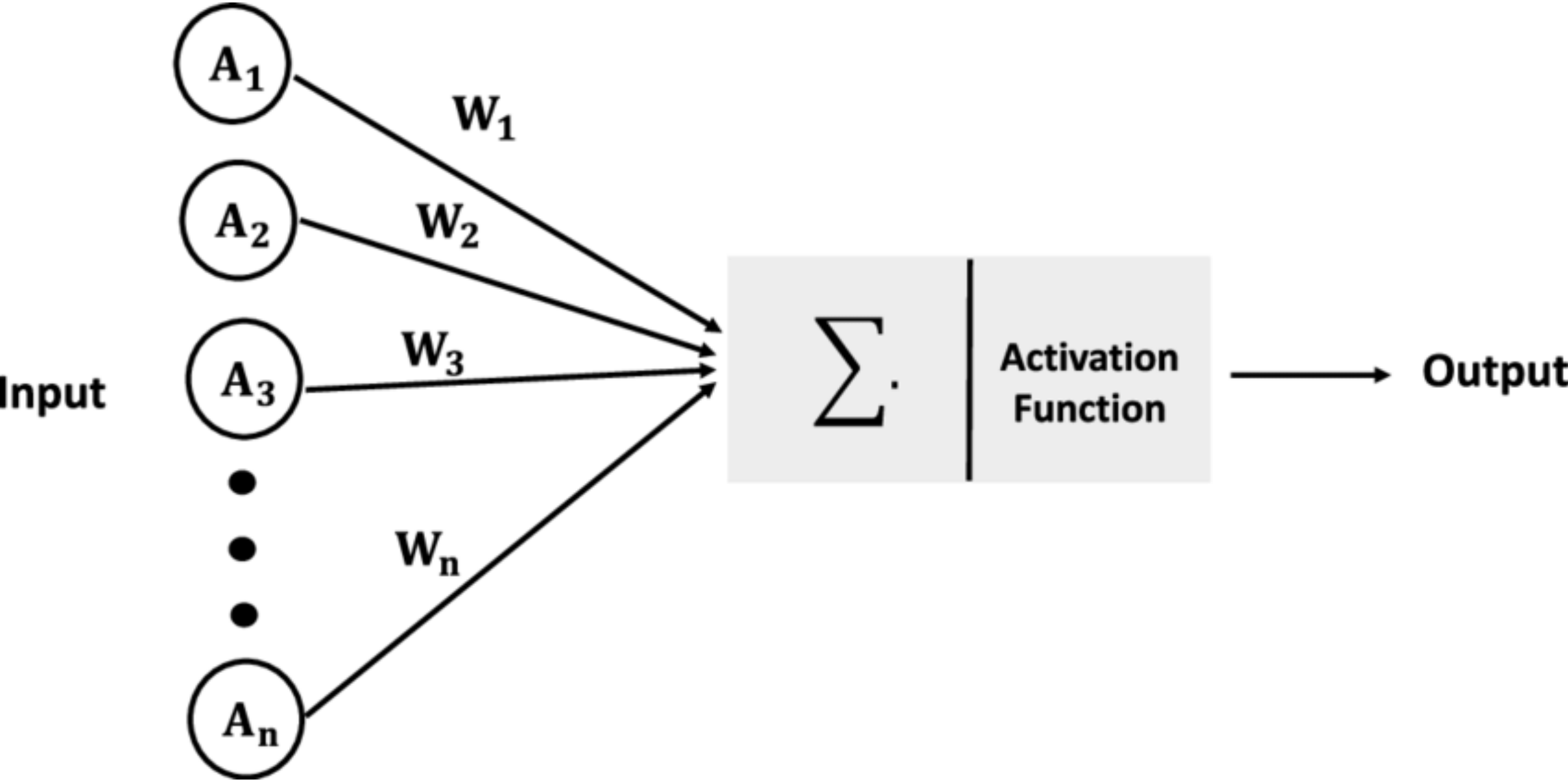
Encoding Tensor (Sophia):
[1,0,1,0,1]
Label Tensor (Sophia):
[aa: 0, […], op:1, […], so:1, […]]



Input:
[1,1,0,0,0]
Output
[aa: 0, […], ma:0.7, […], rc:0.9, […]]

[ma,rc,ce]
↓
Marcel


Input layer
Hidden layers
Output layer
i
h_1
h_2
h_n
o
Input 1
Input 2
Input n
Output 1
Output n


A_1
A_2
A_3
A_n
W_1
W_2
W_3
W_n
Input
$\sum$.
Activation Function
Output